Biweekly Report 4

Adding Gillespie stochastic simulations and validating on small networks

Raunak Narwal

Department of Mathematical Sciences

Indian Institute of Science Education and Research, Mohali, 130406, Punjab

October 24 to November 7, 2025

Gillespie Stochastic Simulation Algorithm (SSA)

The Gillespie Stochastic Simulation Algorithm (SSA) is a computational method which is used to simulate the time evolution of chemical reacting systems, mainly when dealing with small populations of molecules where stochastic effects are significant. Unlike deterministic approaches that use ODEs to model average behavior, Gillespie's SSA captures the inherent randomness of reaction events at the molecular level.

Our dynamic comparison framework was originally designed to parsea and compare deterministic reaction networks in KGML format using ODEs , it lacked a stochastic simulation component. To address this, we have integrated the Gillespie SSA into our existing framework, allowing us to simulate and compare the dynamics of biochemical networks under stochastic conditions. This enhancement enables us to capture the variability and noise present in real biological systems, providing a more comprehensive comparison between different reaction networks. In biochemical kinetics, deterministic ODEs model the average concentration dynamics of species over time:

$$\frac{d[\mathbf{X}]}{dt} = f([\mathbf{X}], \mathbf{k})$$

where $[\mathbf{X}]$ are concentrations and $\mathbf{k}$ are rate constants.

Gillespie SSA models the discrete molecular events that occur probabilistically:

$$a_r = c_r \prod_i n_i^{\nu_{ri}}$$

where a_r is the propensity of reaction r, c_r is the stochastic rate constant, n_i are molecule counts, and ν_{ri} is the stoichiometry.

These two formulations are theoretically equivalent in the limit of large molecular populations and nicely mixed conditions. However, converting between them requires scaling

1

by Avogadro's number N_A and system volume V which was a lot of conversions with our code. The relationship between deterministic rate constants k_r and stochastic rate constants c_r is given by:

$$c_r = \frac{k_r}{N_A^{\sum_i \nu_{ri}-1} V^{\sum_i \nu_{ri}-1}}$$

where $\sum_i \nu_{ri}$ is the total number of reactant molecules in reaction r. This scaling ensures that the average behavior of the stochastic model matches the deterministic ODEs when molecule counts are large.

Added Components in the Code

SBML parser to support smaller toy networks for example repressilator, futile cycles, predator prey systems and more. This enables us to import any synthethic network directly from BioModels or similar repository.

Gillespie Stochastic Simulation engine that implements a new simulator, based on the standard SSA, but it is extended with MonteCarlo averaging (multiple independent runs), automatic volume estimation (that might be a disadvantage too), diagnostic and logs and optional animations.

Helper function of volume estimation was added to set a physcially consistent and stable system volume, it aimed to produce a target number of events within the window of simulation preventing frozen or explosive dynamics. New simulation controls are added in the CLI. –dynamic gillespie to run the SSA „ –volume V for volume (if no volume is provided the code automatically cimputes it using the estimator, it is better to not provide volume to see comparisons between ODE and SSA), –animate yes for animations of the networks.

Autoscaling Volume

Our first code ran nicely and produced visually appealing results but the concentrations were way off and turns out we did not convert the molecules to concentrations, so we used fixed volumes for conversion for example $1e-15$, but this resulted in zero errors and flat lines of trajectory, clearly something was off. So we used defined an automatic volume estimator, the auto scaling approach dynamically determines a suitable volume to yield a manageable event count and biologically realistic concentrations.

Deterministic Stochastic Convergence

This relationship is a cornerstone of chemical kinetics theory:

$$\lim_{N_A V \to \infty} \frac{n_i}{N_A V} = [X_i]$$

At large V, the Gillespie average becomes identical to the ODE solution.

Results

We tested our Gillespie SSA and our ODE implementations on *map00400* and *map00900*, the results showed similarities in MSE and other error values, at higher mc runs the stochastic trajectories converged closer to the deterministic ones, confirming the theoretical expectation.

```
    "metrics for Gillespie SSA": {
    "dtw_avg": 6.198763329018459,
    "mse_avg": 2.1381405810718794,
    "corr_avg": 0.13299220130476633,
```

```
    }
    "metrics for ODE": {
    "dtw_avg": 6.204106038148442,
    "mse_avg": 2.1387896309983585,
    "corr_avg": 0.13306254705179113,
    }
    "metrics for SSA mc=10": {
    "dtw_avg": 6.1989763329018459,
    "mse_avg": 2.1384581210718794,
    "corr_avg": 0.13199220130476633,
    }
```
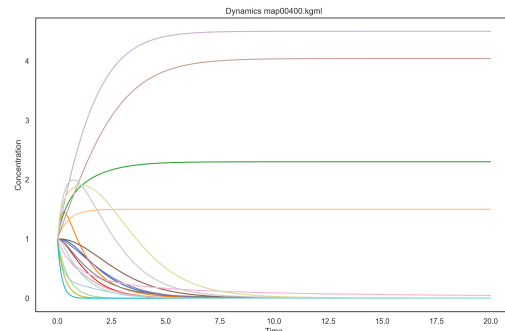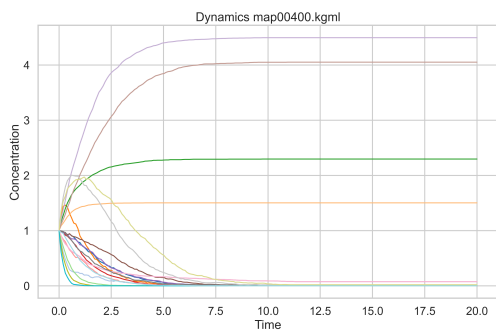

The left one is gillespie SSA and right one is ODE, both the dynamics are very similar, the fluctuations in SSA are due to stochasticity, but overall trends match well.
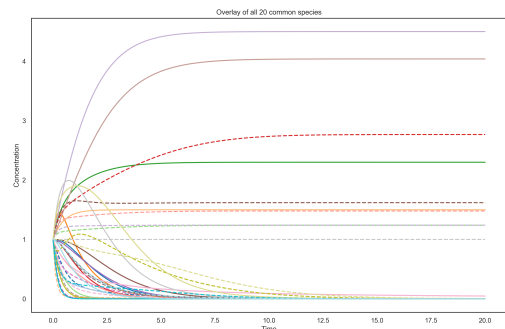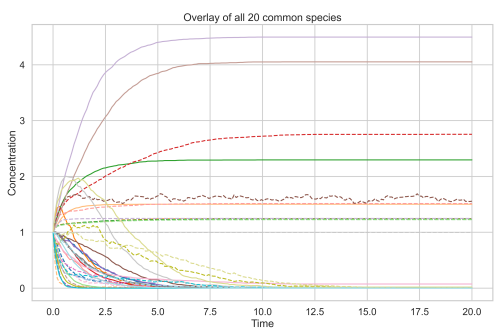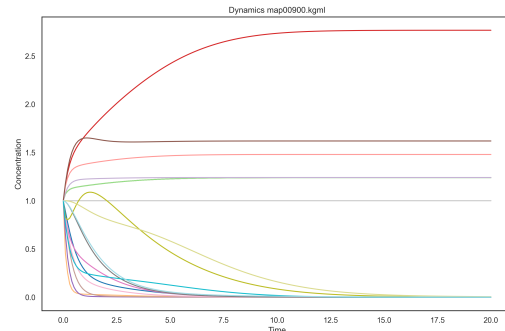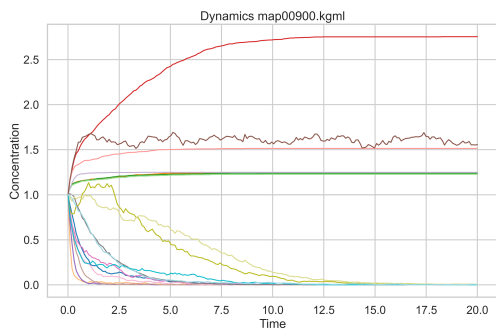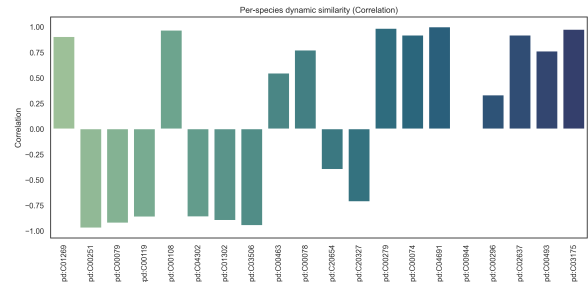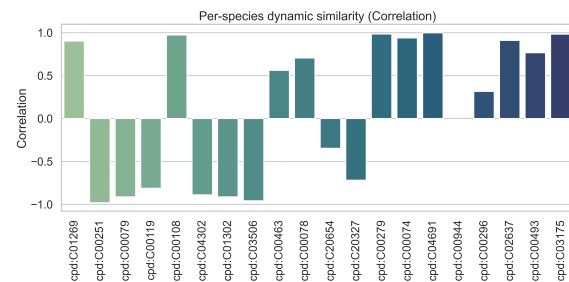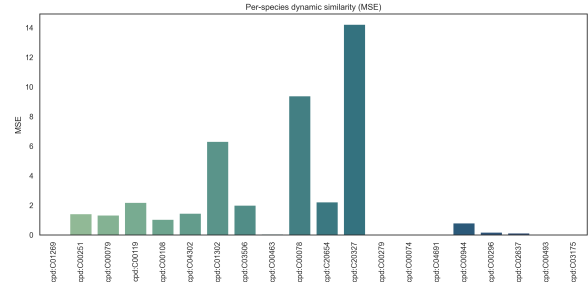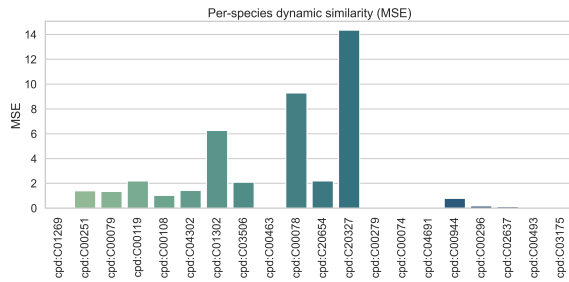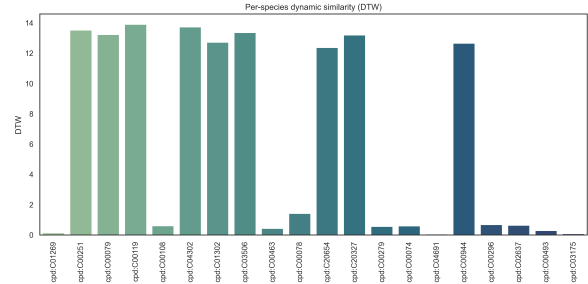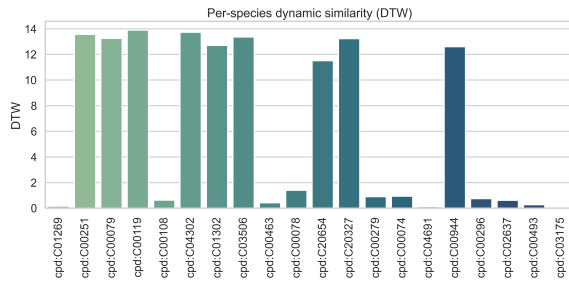
The animation videos are also mostly similar.

Validation on Small Networks

To be continued, it does create lots of errors such as SBML (XML) files dont have a KEGG ID so they have zero species in common which makes dynamic comparison invalid, for topological comparison, we don't yet have a framework to convert SBML files into CSV, so all of these would be done in the next report along with other pending agendas.

Future Work Timeline

Event	Timeline
Biweekly Report 1	21 August
Biweekly Report 1.1	25 August
Mid semester exams	1-2 September
Biweekly Report 2	~~3 September~~ 7 September
Mid semester exams	~~3 September~~ 10 September
Awaiting Feedback	September-October
Mid semester II exams	10-16 October
Biweekly Report 3	~~21 September~~ 24 October
Biweekly Report 4	7 November
End Semester Exams	14-26 November
Biweekly Report 4.1	2 December

Generated: November 8, 2025